



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Detecting anthropogenic influence with a multi-model ensemble

**Citation for published version:**

Gillett, NP, Zwiers, FW, Weaver, AJ, Hegerl, GC, Allen, MR & Stott, PA 2002, 'Detecting anthropogenic influence with a multi-model ensemble', *Geophysical Research Letters*, vol. 29, no. 20, pp. 1-4.  
<https://doi.org/10.1029/2002GL015836>

**Digital Object Identifier (DOI):**

[10.1029/2002GL015836](https://doi.org/10.1029/2002GL015836)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Geophysical Research Letters

**Publisher Rights Statement:**

Published in Geophysical Research Letters by the American Geophysical Union (2002)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Detecting anthropogenic influence with a multi-model ensemble

N. P. Gillett,<sup>1</sup> F. W. Zwiers,<sup>2</sup> A. J. Weaver,<sup>1</sup> G. C. Hegerl,<sup>3</sup> M. R. Allen,<sup>4</sup> and P. A. Stott<sup>5</sup>

Received 8 July 2002; revised 30 August 2002; accepted 3 September 2002; published 23 October 2002.

[1] Averaging results from multiple models has previously been found to improve estimates of the climatology and seasonal predictions of atmospheric variables. Here we describe how a multi-model mean of the simulated response to greenhouse gas and sulphate aerosol changes may be used to detect anthropogenic influence on surface temperature. The scaling factor on a combined greenhouse gas and sulphate aerosol response pattern is estimated using a five model ensemble, and is found to be similar to that estimated using individual models, with similar uncertainties. When applied to the simultaneous detection of separate greenhouse gas and sulphate aerosol responses, the multi-model method indicates a closer consistency between the observations and simulated responses, with reduced uncertainties. This improvement is at least in part due to the larger ensemble sizes and increased length of control integration available when data from multiple models are combined. *INDEX TERMS*: 1610 Global Change: Atmosphere (0315, 0325); 1694 Global Change: Instruments and techniques; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation. **Citation**: Gillett, N. P., F. W. Zwiers, A. J. Weaver, G. C. Hegerl, M. R. Allen, and P. A. Stott, Detecting anthropogenic influence with a multi-model ensemble, *Geophys. Res. Lett.*, 29(20), 1970, doi:10.1029/2002GL015836, 2002.

## 1. Introduction

[2] The IPCC Third Assessment Report [Mitchell *et al.*, 2001] describes the application of “optimal fingerprinting” techniques [Hasselmann, 1997] to the detection of a combined greenhouse gas plus sulphate aerosol (GS) response over the past 50 years. The response patterns simulated by seven climate models were found to be detectable in observations of surface temperature; the amplitudes of these patterns in the observations were found to be inconsistent with zero. However, there were considerable differences in the magnitude of the response between the models, with some simulating a response consistent with that observed, while others predicted a response significantly larger than that observed. When these techniques were used to estimate the amplitudes of the greenhouse gas (G) and sulphate aerosol (S) response patterns separately, these inter-model differences became larger, with simultaneous detection of G

and S possible with some models, but not with others. These results from multiple models are synthesized only qualitatively by Mitchell *et al.* [2001]. Here we suggest a method for doing so more quantitatively.

[3] Lambert and Boer [2001] compared coupled model climatologies of surface air temperature with observations using the CMIP ensemble. They found that overall the mean climatology of the models matched the observations better than that of any individual model. Similarly in seasonal forecasting, Krishnamurti *et al.* [1999] and Kharin and Zwiers [2002] argued that a weighted sum of multiple model predictions performs better than predictions using an individual model. These results might be explained if each model has independent errors, each giving different characteristic biases in model output. Thus, much as we can reduce the effects of initial condition uncertainty by averaging over an ensemble of integrations with perturbed initial conditions, so we might also account for model uncertainty by averaging over multiple models. Such an argument is also likely to apply to the anthropogenically-forced responses of these models. Thus here we describe how a mean of the response patterns of five climate models (HadCM2, HadCM3, ECHAM3, CGCM1 and CGCM2) may be used to detect greenhouse gas and sulphate aerosol influences in surface air temperature.

[4] In one standard approach to the detection of anthropogenic influence on surface temperature [Allen *et al.*, 2002], signal-to-noise optimised observations are regressed against a modelled response pattern, using a total least squares fit. Climate model output enters the analysis at three points. First, output from integrations of the model with prescribed time-varying forcings is used to derive the signal patterns of climate response. Second, output from a control integration is used to estimate the autocovariance matrix representing internal variability. This covariance matrix is used to derive the EOF basis used for truncation of the signal patterns and signal-to-noise optimisation. Third, an independent section of control data is used to estimate the uncertainty in the derived regression coefficients. In this study, output from multiple models is used in all three stages.

## 2. Method

[5] Kharin and Zwiers [2002] describe the use of multiple models for seasonal predictions, where 10 years of data were available to compare with observations (and therefore 10 verifying periods), and found that in many cases the quantity of verifying data was too limited to use anything other than a simple multi-model mean. In this case, we have essentially only one verifying period (the late twentieth century); thus any more complex weighted combination of model response patterns would certainly be under-constrained. However, since both internal variability and model

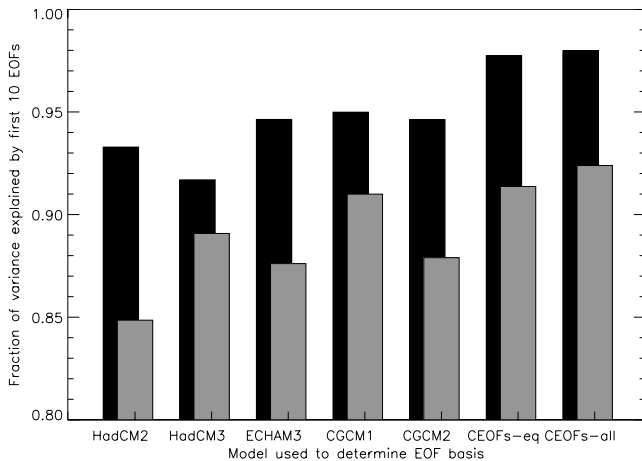
<sup>1</sup>School of Earth and Ocean Sciences, University of Victoria, British Columbia, Canada.

<sup>2</sup>Canadian Centre for Climate Modelling and Analysis, Victoria, British Columbia, Canada.

<sup>3</sup>Duke University, Durham, North Carolina, USA.

<sup>4</sup>Department of Physics, University of Oxford, UK.

<sup>5</sup>Met Office, Hadley Centre, Bracknell, UK.



**Figure 1.** Fraction of variance explained by 10 EOFs derived from control integrations of the models shown in the multi-model 1946–1996 GS response pattern (black bars) and the corresponding decadal mean observations (grey bars). CEOFs-eq refers to common EOFs derived using equal length control segments from each model, and CEOFs-all refers to common EOFs derived using half of each model’s control integration. Total lengths of control integration used to derive the EOFs were 860, 560, 850, 470, 470, 2350 and 3210 years respectively.

error contribute to uncertainty in the model response patterns, it is sensible to give models with large ensembles more weight. How much more depends on the relative importance of internal variability and model error, but for the present analysis we take the simple approach of weighting each model integration equally.

[6] Having derived multi-model response patterns, we also need to derive the EOF basis used in the truncation and signal-to-noise optimisation. The natural approach to representing the mean variability in our multi-model ensemble is to use common EOFs of the control integrations of all the models used. However the control integrations are of unequal lengths; thus we need to decide whether to give equal weights to each model, or equal weights to each control segment. We tried both approaches: first concatenating output from half of each available control integration, and second concatenating equal lengths of control data from each model. The fraction of variance explained by the first 10 EOFs (the truncation chosen by *Allen et al.* [2002]) of each model control integration and in each common EOF basis is shown for the multi-model GS response pattern and the observations in Figure 1. This was derived by squaring the correlation coefficient between the raw and truncated pattern in each case. The common EOF bases explain the highest fraction of variance of both the modelled response pattern and the observations, and this result was found to be robust to the use of a second independent control segment for the calculation of the EOFs in each case. These results also show that the fraction of variance explained by the common EOFs is insensitive to the relative weightings given to each model: the use of all available data gives slightly better results, thus we take this

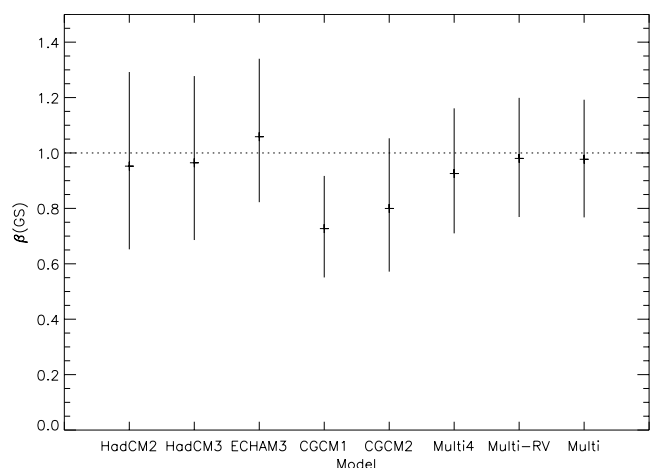
approach hereafter. We also found that the regression coefficients are insensitive to the choice of model weighting used to derive the common EOFs.

[7] The final input taken from the models is an independent section of control data used to estimate the uncertainty ranges on the derived regression coefficients. Following the approach in earlier sections, we concatenated the remaining halves of each model’s control integration for this part of the analysis. The use of equal length control segments from each model was again found to have little influence on the results.

### 3. Results

[8] We started by applying our multi-model approach to the pattern of surface temperature changes simulated in response to historical reconstructions of greenhouse gas and sulphate aerosol changes (GS). We followed the approach of *Allen et al.* [2002], using five decadal means (1946–1996) of surface temperature expressed as anomalies from the 1906–1996 climatology, projected onto T4 spherical harmonics. Observations are updated from those described by *Parker et al.* [1994]. Figure 2 shows the best estimate of the regression coefficient,  $\beta_{GS}$ , for the GS pattern, along with its uncertainties. This is the factor by which the model response must be scaled to best match the observations.

[9] In the multi-model case, this scaling factor is both significantly greater than zero (GS is said to be “detected”), and it is also consistent with one (the modelled response is consistent with observations). All the models also individually detect their GS patterns and, apart from CGCM1, obtain scaling factors consistent with 1 (see also *Allen et al.*



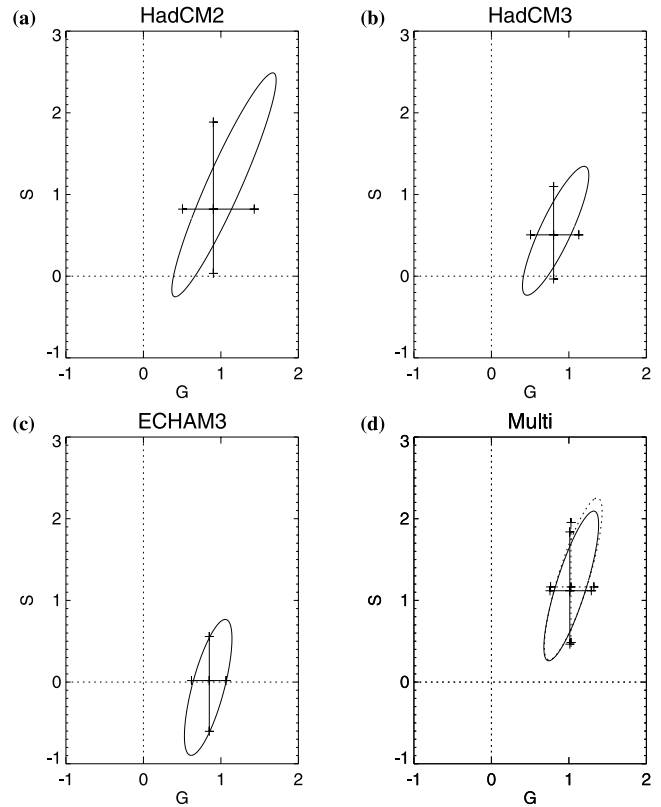
**Figure 2.** Factors,  $\beta_{GS}$ , by which a modelled GS response pattern must be scaled to best match observations. The first five bars represent results derived separately from each model, and the last one the results from a multi-model analysis. 5–95% uncertainty ranges are shown. The bar labelled ‘Multi4’ represents results derived using a restricted four-member, four-model ensemble. The bar labelled ‘Multi-RV’ shows results from the multi-model ensemble where intra-ensemble variability has been used to estimate the uncertainty in the response pattern.

[2002]). Note that uncertainty ranges are not much smaller in the multi-model case than for individual models. This reflects the fact that much of the uncertainty is due to internal variability in the observations. Overall we might conclude that the multi-model approach is a good way to synthesize GS response results from multiple models, but other than that it provides no clear benefits. Restricting the multi-model ensemble to one integration from each of four models (HadCM2, HadCM3, CGCM1 and CGCM2), thereby making results directly comparable to those derived using the four-member HadCM2 and HadCM3 ensembles, did not have a large effect on the results (regression coefficient labelled ‘Multi4’ in Figure 2).

[10] Applying optimal detection techniques to the GS pattern is somewhat restrictive, because the assumption is made that the relative size of the responses to greenhouse gases and sulphate aerosol is correctly simulated by these models. However, by using output from G integrations along with output from GS integrations we may relax this assumption, and estimate the amplitudes of the G and S response patterns separately. To do this, we make the assumption that the response to the combined forcings is the linear sum of the responses to the forcings individually [Haywood *et al.*, 1997]. This approach was applied using HadCM2, ECHAM3 and ECHAM4 by Allen *et al.* [2002], who found that G and S were only both separately detectable using HadCM2.

[11] We used the multi-model approach to separately estimate the amplitudes of the G and S scaling factors in a two-way regression. Figure 3 shows the estimated scaling factors on the G and S response patterns, along with the associated uncertainty ranges for the multi-model case and individually for HadCM2, HadCM3 and ECHAM3 [Allen *et al.*, 2002]. In the multi-model case, the G and S response patterns are separately detected, and they both have amplitudes consistent with one. The only other model in which both G and S signals are detected is HadCM2. Note that CGCM1 and CGCM2 did not have ensembles of greenhouse gas only integrations, so they could not be used on their own to separately estimate G and S amplitudes. Nonetheless, output from these models was still used to estimate the GS response pattern and to contribute to our estimate of control variability.

[12] Figure 3 shows that in the multi-model case the estimated regression coefficients are closer to one and the uncertainties are reduced compared to the single model cases. One might interpret these results as showing that by averaging over multiple models, we have reduced the impact of individual model errors, thereby deriving a less uncertain response pattern closer to that observed. In fact it appears that the main influence arises from using multiple models to improve our estimate of control variability, giving an EOF basis which better resolves the greenhouse gas and sulphate aerosol response patterns. If multi-model control data is used together with single model response patterns, G and S are separately detected for all the models shown in Figure 3. Allen *et al.* [2002] estimate the uncertainty in the model response patterns by assuming that intra-ensemble variability is well-modelled by control variability. If we use a multi-model ensemble, we might no longer expect this assumption to be valid. However, rescaling our estimate of the variance in the response pattern by the ratio of intra-



**Figure 3.** Scaling factors,  $\beta_G$  and  $\beta_S$ , by which modelled G and S response patterns must be scaled to best match observations. Uncertainty bars show one-dimensional 5–95% confidence ranges and the two-dimensional 90% confidence region is shown as an ellipse. Results are based separately on output from (a) HadCM2, (b) HadCM3, (c) ECHAM3, and (d) the multi-model ensemble. Dotted lines in (d) show results for which intra-ensemble variability has been used to estimate the uncertainty in the response pattern.

ensemble variance (the variance calculated over the whole multi-model ensemble) to control variance, thereby implicitly folding an estimate of model uncertainty into our analysis, was found to have little effect on our results (compare ‘Multi’ and ‘Multi-RV’ bars in Figure 2 and solid and dotted ellipses in Figure 3d). It may be thought that the apparent improvement in the multi-model case results purely from the inclusion of models with lower variability (CGCM1, CGCM2) in the analysis. To test this, we re-scaled the control variance of the multi-model ensemble to be equal to that of the highest variance model, HadCM2. This inflated the uncertainty intervals by a factor of 1.3, but detection of G and S was still more certain than using any individual model.

#### 4. Conclusions

[13] Optimal detection techniques were applied using a multi-model ensemble containing five models (HadCM2, HadCM3, ECHAM3, CGCM1 and CGCM2). The method was first applied to the detection of a single GS response pattern in surface temperature. The estimated scaling factor

on the multi-model GS response was found to lie near the centre of the range of those estimated using individual models, with comparable uncertainties. The GS pattern was detected and found to be consistent between models and observations. In this case, the multi-model approach offers a method for synthesizing results from more than one model, but no other clear benefits.

[14] The multi-model approach was then applied to the simultaneous detection of G and S. Both G and S were separately detected, unlike for all the constituent models apart from HadCM2. The estimated scaling factors were closer to one and had smaller uncertainties than those derived using individual models. Thus in this more exacting test of the methodology, the use of multiple models helped to reduce uncertainties in the calculated regression coefficients. This is at least partly because of the larger ensemble sizes and increased volume of control data available when output from multiple models is combined. Internal variability in the observations makes it difficult to determine whether the reduction in model error through averaging over multiple model responses is also important.

[15] **Acknowledgments.** We are grateful to NSERC and CFCAS for CLIVAR funding. PAS was funded by the UK Department of the Environment, Food and Rural Affairs under contract PECD 7/12/37.

## References

- Allen, M. R., et al., Quantifying anthropogenic influence on recent near-surface temperature change, *Surveys in Geophysics*, accepted, 2002.
- Hasselmann, K., Multi-pattern fingerprint method for detection and attribution of climate change, *Clim. Dyn.*, **13**, 601–611, 1997.
- Haywood, J. M., et al., Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations, *Geophys. Res. Lett.*, **24**, 1335–1338, 1997.
- Kharin, V. V., and F. W. Zwiers, Climate predictions with multimodel ensembles, *J. Clim.*, **15**, 793–799, 2002.
- Krishnamurti, T. N., et al., Improved weather and seasonal climate forecasts from multimodel superensemble, *Science*, **285**, 1548–1550, 1999.
- Lambert, S. J., and G. J. Boer, CMIP1 evaluation and intercomparison of coupled climate models, *Clim. Dyn.*, **17**, 83–106, 2001.
- Mitchell, J. F. B., et al. Detection of climate change and attribution of causes, edited by J. T. Houghton, et al., *Climate change 2001, The scientific basis*, chapter 12, pp. 695–738. Cambridge Univ. Press, 2001.
- Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, Interdecadal changes of surface temperature since the late nineteenth century, *J. Geophys. Res.*, **99**, 14373–14399, 1994.

---

N. P. Gillett, A. J. Weaver, and F. W. Zwiers, SEOS, University of Victoria, PO Box 3055, Victoria, BC, V8W 3P6, Canada. (gillett@ocean.seos.uvic.ca)

M. R. Allen, AOPP, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, UK.

G. C. Hegerl, Dept. of Earth and Ocean Sciences, Nicholas School for the Environment, Duke University, Box 90227, Durham, NC 27708, USA.

P. A. Stott, Met Office, Hadley Centre, London Road, Bracknell, Berkshire, RG12 2SY, UK.